Optimal Capacity Allocation and Performance Analysis of Communication Networks

著者（英）	Jun Wu, Wuyi Yue, Shouyang Wang
journal or publication title	Memoirs of Konan University. Science and engineering series
volume	52
number	1
page range	105-124
year	2005-07-31
URL	http://doi.org/10.14990/00000092

Mem. Konan Univ., Sci. & Eng. Ser., **52** (1) 105 ~ 124 (2005)

Optimal Capacity Allocation and Performance Analysis of Communication Networks

[a,c]Jun Wu, [b]Wuyi Yue and [c]Shouyang Wang

[a]Institute of Intelligent Information and Communications Technology
Konan University, 8-9-1 Okamoto, Kobe 658-8501 Japan, e-mail: wujun@iss.ac.cn
[b]Department of Information Science and Systems Engineering
Konan University, 8-9-1 Okamoto, Kobe 658-8501 Japan, e-mail: yue@konan-u.ac.jp
[c]Academy of Mathematics and Systems Science
Chinese Academy of Sciences, Beijing 100080 China, e-mail: sywang@amss.ac.cn

(Received April 14, 2005)

Abstract

In this paper, we present a stochastic model for optimizing capacity allocation in Communication Networks (CNs) with a loss rate constraint, a profit shortfall risk and a penalty cost. The model is proposed for offline traffic engineering optimization, taking a centralized view of bandwidth allocation, performance control and risk of profit shortfall. First, we present an analysis of the capacity allocation problem to derive the optimal bandwidth capacity with a loss rate constraint. Due to the random traffic load, there exists a risk of CN profit shortfall. We use a mean-variance approach to analyze the risk averseness in CNs and we also derive the optimal bandwidth capacity with risk averseness. Second, we present an analysis of the capacity allocation problem to derive the optimal bandwidth capacity with a penalty cost. Finally, numerical results are shown to reveal the impacts of loss rate constraint, risk averseness and penalty cost on the CNs' performance. The results presented in this paper reveal insights for CNs' design and planning.

Keywords: Communication networks, stochastic traffic engineering, loss rate constraint, penalty cost, optimization, risk averseness.

1 Introduction

Traffic engineering in Communication Networks (CNs) is a process of controlling traffic demand in a network so as to optimize resource utilization and network performance [1], [2]. There are two forms of traffic engineering: online planning and offline planning. Online traffic engineering focuses on instantaneous network states and individual connections. While offline traffic engineering simultaneously examines each channel's resource constraints and the requirements of each Local Service Provider (LSP) to provide global calculations and solutions for the CNs in a centralized view. Traffic engineering has greatly improved network utilization and performance by using the emerging technologies, such as Multi-Protocol Label Switching (MPLS) and Optical Channel Trails (OCT) [3], [4].

In previous works, the offline optimization problem was formulated as a deterministic Multi-Commodity Flow (MCF) model, where demand of each channel was given as a fixed quantity [5], [6]. The objective usually aims to optimize the network total revenue from serving traffic demand. In the deterministic MCF model, network revenue was assumed to be a linearly increasing function of the amount of bandwidth provisioned up to the capacity, where all traffic demand was satisfied.

The deterministic approach in the above is inadequate in a case of random demand. When the demand is random, we cannot know the traffic load exactly, so there exists a risk of over-provisioned bandwidth capacity or less-provisioned bandwidth capacity, either is bad for the network performance. Due to the random demand, the network profit is uncertain, which is dependent on the underlying demand with a specified probability function. So, there exists a risk of network profit shortfall, i.e., the network cannot obtain the expected profit confessedly.

Recently, there were some works concerning with the stochastic traffic engineering. Mitra and Wang presented a stochastic traffic engineering framework for optimizing bandwidth provisioning and path selection in CNs [7]. The objective was to maximize revenue from serving demand in a two-tier market in telecommunication industry. Mitra and Wang also developed an optimization framework for the network service provider to manage profit in a two-tier market [8]. They investigated the impact of network size and risk averseness on the bandwidth management. Mitra and Wang furthered their studies in [7], [8] and developed the efficient frontier of mean revenue and revenue risk [9].

But all the above researches on stochastic traffic engineering did not investigate the network performance control, which is an important aspect in traffic engineering planning. Motivated by this, in this paper, we develop a stochastic traffic engineering optimization model focusing on both the random traffic demand and the network performance optimization, such as loss rate and network risk management. First, based on the model presented in [7], we introduce a probabilistic constraint to guarantee the loss rate within a specified level. Then, by using mean-variance approach, we study the impact of risk averseness on the network profit function. Finally, we introduce a penalty cost in the optimization model for network bandwidth allocation. Whenever there is unsatisfied traffic demand with the limitation of network bandwidth, a penalty function will be added in the objective function.

The rest of this paper is organized as follows. In Section 2, we present the system model that we consider in this paper and present the notations and preliminaries. In Section 3, we formulate the optimization model and derive the optimal bandwidth capacity with a loss rate constraint. In Section 4, we analyze the network profit shortfall risk by using a mean-variance approach. In Section 5, we formulate the optimization model and derive the optimal bandwidth capacity with penalty cost. In Section 6, we give some numerical results to show the impacts of loss rate constraint, risk averseness and penalty cost on the network performance. Conclusions are given in Section 7.

2 System Model

A Communication Network (CN) is formulated as a collection of nodes and links, that should derive its revenue by delivering traffic load to and from its users. For unit bandwidth capacity allocated to the network, a unit cost will be charged. For unsatisfied traffic demand with the limitation of network bandwidth, a penalty cost will be added in the objective function. The objective of this system is to maximize the expected profit of the network. To guarantee the network performance, we introduce and present analyses for the loss rate constraint, the risk of profit shortfall and the penalty cost.

Let (N, L) denote a CN composed of nodes v_i ($v_i \in N$, $1 \le i \le N$) and links l ($l \in L$), where N is the set of all nodes and L is the set of all links. Let V denote the set of all node pairs, $v \in V$ denote an arbitrary node pair where $v = (v_i, v_j)$ and v_i, $v_j \in N$. Let C_l denote the maximal bandwidth capacity of link l, $R(v)$ denote an admissible route set for $v \in V$, ξ_s ($s \in R(v)$) denote the amount of capacity provisioned on route s, D_v ($v \in V$) denote the traffic load between node pair $v \in V$, b_v ($v \in V$) denote the amount of bandwidth capacity provisioned between node pair v, which can be routed on one or more routes, then $b_v = \Sigma_{s \in R(v)} \xi_s$.

In this paper, we consider the CN to be a whole system. We let b denote the amount of bandwidth capacity provisioned in the CN, then we have $b = \Sigma_{v \in V} b_v$. Let D denote the traffic demand in the whole CN, then we have $D = \Sigma_{v \in V} D_v$, which is characterized by a random distribution with its probability density function $f(x)$ and cumulative distribution function $F(x)$. $b \wedge D$ is the actual traffic load transmitted in the CN, where $\wedge$ represents the choice of the smaller value between b and D. Let r denote the unit revenue by serving the traffic demand, so the total revenue of the CN is $r \times (b \wedge D)$. Let c denote the unit cost for unit bandwidth capacity allocated in the CN, so the total cost is $c \times b$. Let p denote the penalty cost for each unsatisfied traffic demand with the limitation of network bandwidth, so the total penalty cost is $p\times(D-b)^+$, where "+" represents the choice of the positive part of $(D-b)$. Let $P(b\geq\delta D)\geq 1-\epsilon$ denote the loss rate constraint with δ and ϵ as the parameters. In the loss rate constraint, δ is the percentage of satisfied users and $1-\delta$ is the loss rate. As δ increases, the loss rate $1-\delta$ deceases. The higher δ is, the better the network performance is. $1-\epsilon$ is the confidence level, which represents the probability of $b\geq\delta D$. As ϵ decreases, the confidence level $1-\epsilon$ increases. A higher confidence level $1-\epsilon$ guarantees a higher probability of achieving a better network performance.

To avoid unrealistic and trivial cases, we make the following assumptions:

(1) Probability density function $f(x)$ of the random traffic load is $f(x)\geq 0$.

(2) Cumulative distribution function of the random traffic load $F(x)$ is strictly increasing in x.

(3) Traffic load D in the CN is assumed to be positive, i.e., $D>0$.

(4) Bandwidth capacity b provisioned in the CN is assumed to be positive, i.e., $b>0$.

(5) Maximal capacity C_{max} that can be allocated in the CN is assumed to be positive, i.e., $C_{max}>0$.

(6) System parameters unit revenue r, penalty cost p and unit cost c satisfy $r>p>0$, $r>c>0$.

(7) Loss rate constraint parameter δ satisfies $0\leq\delta\leq 1$.

(8) Confidence level $1-\epsilon$ satisfies $0\leq 1-\epsilon\leq 1$.

3 Optimal Capacity Allocation with Loss Rate Constraint

In this section, we formulate the model for the network bandwidth allocation problem and derive the optimal bandwidth capacity with a loss rate constraint.

Let $\pi(b, D)$ denote the random profit function by transmitting messages in the network, namely,

$$\pi(b, D) = r(b\wedge D) - cb. \tag{1}$$

Let $\Pi(b, D)$ denote the mean profit function as follows:

$$\Pi(b, D) = E[\pi(b, D)] = r\int_0^b xf(x)dx + rb\int_b^{+\infty} f(x)dx - cb. \tag{2}$$

The objective function of the system is given by

$$\Pi^* = \max_{b>0}\{\Pi(b, D)\}, \tag{3}$$

subject to

$$P(b \geq \delta D) \geq 1-\epsilon \tag{4}$$

and

$$b \leq C_{max} \tag{5}$$

where Π^* is the optimal profit function.

There are two cases represented in Eq. (2). The first term in Eq. (2) is the case of $D \leq b$, the second term in Eq. (2) is the case of $D > b$. Eq. (4) is the loss rate constraint. It means that the capacity provisioned in the CN guarantees a loss rate less than $1-\delta$ with the confidence level larger than $1-\epsilon$. The loss rate constraint enables us to control the network performance by properly setting the system parameters.

With the above preparation, we can derive the optimal bandwidth capacity that should be allocated in the CN. First, we analyze the property of the mean profit function $\Pi(b, D)$ without any constraints.

The first order derivative of $\Pi(b, D)$ of Eq. (2) with respect to b is given as follows:

$$\frac{d\Pi(b,D)}{db} = (r-c) - rF(b). \tag{6}$$

The second order derivative of $\Pi(b, D)$ of Eq. (2) with respect to b is given as follows:

$$\frac{d^2\Pi(b,D)}{db^2} = -rf(b). \tag{7}$$

From the assumptions in Section 2, we know that $f(b) \geq 0$, $r > 0$, hence,

$$\frac{d^2\Pi(b,D)}{db^2} \leq 0. \tag{8}$$

Therefore, we can say that $\Pi(b, D)$ is a concave function of b. So, the optimal bandwidth capacity that allocated in the CN is given by

$$F^{-1} = \left(\frac{r-c}{r}\right) \tag{9}$$

where $F^{-1}(\cdot)$ is the inverse function of $F(\cdot)$.

Next we analyze the loss rate constraint. Note that the loss rate constraint is equivalent to

$$P(b \geq \delta D) = P\left(D \leq \frac{b}{\delta}\right) \geq 1-\epsilon. \tag{10}$$

By the definition of cumulative distribution function, Eq. (10) becomes

$$P\left(D \leq \frac{b}{\delta}\right) = \int_0^{\frac{b}{\delta}} f(x)dx = F\left(\frac{b}{\delta}\right) \geq 1-\epsilon. \tag{11}$$

So the loss rate constraint is equivalent to

$$b \in \left[\delta F^{-1}(1-\epsilon), +\infty\right]. \tag{12}$$

Thus, the optimal capacity for the CN bandwidth allocation with loss rate constraint is given as follows:

$$b^* = F^{-1}\left(\frac{r-c}{r}\right) \vee \delta F^{-1}(1-\epsilon) \tag{13}$$

where b^* is the optimal bandwidth capacity, $\vee$ represents the choice of the larger value between $F^{-1}(\frac{r-c}{r})$ and $\delta F^{-1}(1-\epsilon)$.

From Eq. (13), we know that:

(1) if $F^{-1}(\frac{r-c}{r}) \geq \delta F^{-1}(1-\epsilon)$, i.e., the optimal bandwidth capacity satisfies the loss rate constraint, then we can obtain the optimal profit as desired.

(2) if $F^{-1}(\frac{r-c}{r}) < \delta F^{-1}(1-\epsilon)$, i.e., the optimal bandwidth capacity does not satisfy the loss rate constraint, then we cannot obtain the optimal profit. In this case, the maximal profit that we can obtain is $\Pi(F^{-1}(1-\epsilon), D)$.

The critical condition is given as follows:

$$F^{-1}\left(\frac{r-c}{r}\right) = \delta F^{-1}(1-\epsilon). \tag{14}$$

It implies that we can properly set the parameters in Eq. (14) to make a tradeoff between the optimal profit and the network performance.

Finally, if we consider the maximal capacity constraint, then the optimal bandwidth capacity for the network is given as follows:

$$\left[F^{-1}\left(\frac{r-c}{r}\right) \vee \delta F^{-1}(1-\epsilon)\right] \wedge C_{max}. \tag{15}$$

4 Risk Analysis of the Communication Networks

The mean-variance analysis, which was first introduced by Markowitz [10], had been a standard tool in risk management [11]. It involves a systematic tradeoff between the expected return and the specified risk measure. It uses a parameter α ($0.0 \leq \alpha \leq 1.0$) to characterize the risk averseness, which is a quantitative balance between the mean profit and the risk of its shortfall. $\alpha = 0.0$ is the special case of maximizing the mean profit function only. $\alpha = 1.0$ is the special case of most averse to risk. When α increases from 0.0 to 1.0, it indicates the willingness to sacrifice the mean profit to avoid the risk of its variance. Note that, for any given α, the solution is optimal in the sense that we cannot improve the mean profit without increasing the risk, or reduce the risk without decreasing the mean profit [11].

Due to the random traffic load, the profit is also uncertain and is dependent on the distribution of the traffic load. So, in many cases, the optimal profit cannot be obtained as desired. Based on this, we define the risk as the deviation from the optimal profit.

From Section 3, we know that the random profit function of the CN is given by Eq. (1), and the mean profit function is given by Eq. (2). By using the method of integral by parts, Eq. (2) can be obtained as follows:

$$E[\pi(b,D)] = (r-c)b - r\int_0^b F(x)dx. \tag{16}$$

The variance profit function can be obtained as follows:

$$Var[\pi(b,D)] = E\left[(\pi(b,D))^2\right] - \left(E[\pi(b,D)]\right)^2. \tag{17}$$

By using the definition of expectation and method of integral by parts, the first term in Eq. (17) is given as follows:

$$\begin{aligned} E\left[(\pi(b,D))^2\right] &= E\left[r^2(b\wedge D)^2 + c^2b^2 - 2rcb(b\wedge D)\right] \\ &= r^2\int_0^b x^2 f(x)dx - 2rcb\int_0^b xf(x)dx + \left(r^2b^2 - 2rcb^2\right)\int_b^{+\infty} f(x)dx + c^2b^2 \\ &= \left(r^2+c^2\right)b^2 - 2r^2\int_0^b xF(x)dx + 2rcb\int_0^b F(x)dx. \end{aligned} \tag{18}$$

With a similar method, the second term in Eq. (17) is given as follows:

$$\begin{aligned} \left(E[\pi(b,D)]\right)^2 &= r^2\left(\int_0^b xf(x)dx\right)^2 + r^2b^2\left(\int_b^{+\infty} f(x)dx\right)^2 + 2r^2b\int_0^b xf(x)dx\int_b^{+\infty} f(x)dx \\ &\quad + c^2b^2 - 2rcb\int_0^b xf(x)dx - 2rcb^2\int_b^{+\infty} f(x)dx \\ &= (r-c)^2b^2 - 2rb(r-c)\int_0^b F(x)dx + r^2\left(\int_0^b F(x)dx\right)^2. \end{aligned} \tag{19}$$

Substituting Eqs. (18) and (19) into Eq. (17), we can obtain that

$$Var[\pi(b,D)] = 2r^2b\int_0^b F(x)dx - 2r^2\int_0^b xF(x)dx + 2rcb^2 - r^2\left(\int_0^b F(x)dx\right)^2. \tag{20}$$

Based on the above results, we can use mean-variance approach to study the risk of profit shortfall. The objective function, which is denoted by Φ^*, by using the mean-variance analysis is given as follows:

$$\Phi^* = \max_{b>0}\left\{E[\pi(b,D)] - \alpha Var[\pi(b,D)]\right\} \tag{21}$$

where b is the bandwidth capacity, α is the risk averseness parameter, $\pi(b,D)$ is the random profit, $E[\pi(b,D)]$ is the mean of the random profit, and $Var[\pi(b,D)]$ is the variance of the random profit. By substituting Eqs. (16) and (20) into Eq. (21), we get

$$\begin{aligned} \Phi^* = \max_{b>0}\Bigg\{&(r-c)b - r\int_0^b F(x)dx - \alpha\Bigg[2rcb^2 - 2r^2\int_0^b xF(x)dx \\ &- r^2\left(\int_0^b F(x)dx\right)^2 + 2r^2b\int_0^b F(x)dx\Bigg]\Bigg\}. \end{aligned} \tag{22}$$

Next we analyze the properties of the objective function Φ^* given by Eq. (22). From Section 3, we have known that the first term of Eq. (21), the mean function, is a concave function of bandwidth b, so we can focus on the second term of Eq. (21), the variance function.

The first order derivative of $Var[\pi(b,D)]$ of Eq. (20) with respect to b is given as follows:

$$\frac{dVar[\pi(b,D)]}{db} = 4rbc + 2r^2[1-F(b)]\int_0^b F(x)dx. \tag{23}$$

From the assumptions in Section 2, we know that $r>c>0$, $b>0$, hence,

$$4rbc + 2r^2[1-F(b)]\int_0^b F(x)dx > 0.$$

So, the variance function is increasing in b. It means that the risk of profit shortfall increases as the bandwidth capacity increases.

From the above analysis on the properties of mean function and variance function, we know that the mean function $E[\pi(b,D)]$ increases in the interval $b \in (0, F^{-1}(\frac{r-c}{r})]$ and decreases in the interval $b \in [F^{-1}(\frac{r-c}{r}), +\infty)$, the variance function $Var[\pi(b,D)]$ increases in the interval $b \in (0,+\infty)$. When b increases in $(0, F^{-1}(\frac{r-c}{r})]$, both the mean function $E[\pi(b,D)]$ and the variance function $Var[\pi(b,D)]$ will increase. When b increases in $[F^{-1}(\frac{r-c}{r}), +\infty)$, the mean function $E[\pi(b,D)]$ will decrease, while the variance function $Var[\pi(b,D)]$ will keep increasing. Thus, the optimal solution is obviously within the interval $b \in (0, F^{-1}(\frac{r-c}{r})]$.

The above result is quite different from the case without risk averseness. When we analyze the model with risk averseness, we obtain a set $b \in (0, F^{-1}(\frac{r-c}{r})]$ for optimal solutions. The set for optimal solutions means that if we want to improve the mean profit of the CN, then the risk of the profit shortfall of the CN will increase at the same time. Or if we want to reduce the risk of the CN profit shortfall, then the mean profit of the CN will decrease at the same time.

The results presented in this section imply that there are alternative tradeoff solutions for network design: if you want to get more mean profit, then you will bear more risk of its shortfall, if you want to bear less risk, then you will get a lower mean profit.

5 Optimal Capacity Allocation and Risk Analysis with Penalty Cost

Whenever there is unsatisfied traffic demand with the limitation of network bandwidth, a penalty function should be considered in the objective function. Based on the model presented in Section 3, in this section, we derive the optimal bandwidth capacity with a penalty cost and study the risk averseness to evaluate the system performance.

5.1 Optimal Capacity Allocation with Penalty Cost

In this subsection, we derive the optimal bandwidth capacity with a penalty cost for network bandwidth allocation to evaluate the system performance.

Let $\pi_p(b,D)$ denote the random profit function with a penalty cost, namely,

$$\pi_p(b, D) = r(b \wedge D) - p(D-b)^+ - cb. \tag{24}$$

Let $\Pi_p(b, D)$ denote the mean profit function with a penalty cost as follows:

$$\begin{aligned}\Pi_p(b, D) &= E\left[\pi_p(b, D)\right] \\ &= r\int_0^b xf(x)dx + rb\int_b^{+\infty} f(x)dx - p\int_b^{+\infty}(x-b)f(x)dx - cb.\end{aligned} \tag{25}$$

The objective function of the system is given by

$$\Pi_p^* = \max_{b>0}\left\{\Pi_p(b, D)\right\} \tag{26}$$

where Π_p^* is the optimal profit function.

With the above preparation, we can derive that the optimal bandwidth capacity should be allocated in the CN by analyzing the property of the mean profit function $\Pi_p(b, D)$.

The first order derivative of $\Pi_p(b, D)$ of Eq. (25) with respect to b is given as follows:

$$\frac{d\Pi_p(b, D)}{db} = (r+p-c) - (r+p)F(b). \tag{27}$$

The second order derivative of $\Pi_p(b, D)$ of Eq. (25) with respect to b is given as follows:

$$\frac{d^2\Pi_p(b, D)}{db^2} = -(r+p)f(b). \tag{28}$$

From the assumptions in Section 2, we know that $f(b) \geq 0$, $r>p>0$, hence,

$$\frac{d^2\Pi_p(b, D)}{db^2} \leq 0. \tag{29}$$

Therefore, we can say that $\Pi_p(b, D)$ is a concave function of b. So, the optimal bandwidth capacity that is allocated to the CNs is

$$F^{-1}\left(\frac{r+p-c}{r+p}\right). \tag{30}$$

If we consider the loss rate constraint presented in Section 3, then the optimal bandwidth capacity with loss rate constraint is

$$F^{-1}\left(\frac{r+p-c}{r+p}\right) \vee \delta F^{-1}(1-\epsilon). \tag{31}$$

From Eq. (31), we know that:

(1) if $F^{-1}(\frac{r+p-c}{r+p}) \geq \delta F^{-1}(1-\epsilon)$, i.e., the optimal bandwidth capacity satisfies the loss rate constraint, then we can obtain the optimal profit as desired.

(2) if $F^{-1}(\frac{r+p-c}{r+p}) < \delta F^{-1}(1-\epsilon)$, i.e., the optimal bandwidth capacity does not satisfy the loss rate constraint, then we cannot obtain the optimal profit. In this case, the maximal profit that we can obtain is $\Pi_p(F^{-1}(1-\epsilon), D)$.

The critical condition is given as follows:

$$F^{-1}\left(\frac{r+p-c}{r+p}\right)=\delta F^{-1}(1-\epsilon). \tag{32}$$

It implies that we can properly set the parameters in Eq. (32) to make a tradeoff between the optimal profit and the network performance.

Finally, if we consider the maximal capacity constraint, then the optimal bandwidth capacity for the network is given as follows:

$$\left[F^{-1}\left(\frac{r+p-c}{r+p}\right)\vee \delta F^{-1}(1-\epsilon)\right]\wedge C_{max}. \tag{33}$$

5.2 Risk Analysis with Penalty Cost

From Subsection 5.1, we know that the random profit function of the CN is given by Eq. (24), and the mean profit function of the CN is given by Eq. (25). By using the method of integral by parts, Eq. (25) can be obtained as follows:

$$E[\pi_p(b,D)]=-(r+p)\int_0^b F(x)dx-p\int_0^{+\infty} xf(x)dx+(r+p-c)b. \tag{34}$$

The variance profit function can be obtained as follows:

$$Var[\pi_p(b,D)]=E\left[(\pi_p(b,D))^2\right]-\left(E[\pi_p(b,D)]\right)^2. \tag{35}$$

By using the mean-variance approach to study the risk of profit shortfall, the objective function, which is denoted by Ω^*, is given as follows:

$$\Omega^*=\max_{b>0}\left\{E[\pi_p(b,D)]-\alpha Var[\pi_p(b,D)]\right\} \tag{36}$$

where α is the risk averseness parameter, $\pi_p(b,D)$ is the random profit function given by Eq. (24), $E[\pi_p(b,D)]$ is the mean profit function given by Eq. (34), and $Var[\pi_p(b,D)]$ is the variance function given by Eq. (35).

Next, we use the assumption of the traffic demand in the whole CN to study the risk averseness in the network profit shortfall. We consider a fully distributed communication network, where the traffic demand offered to the whole CN forms a Poisson process with arrival rate $\lambda>0$, since in most of the system models in CNs, the arrival process of traffic demand can be assumed to form a Poisson process.

The interarrival times are exponentially distributed with rate λ. Let X be a random variable representing the time between successive demand arrivals in the Poisson process, then we have the probability distribution function $F_X(x)$ and the probability density function $f_X(x)$ of X as follows:

$$F_X(x)=\begin{cases}1-e^{-\lambda x}, & x>0\\ 0, & x\leq 0,\end{cases} \tag{37}$$

$$f_X(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0\\ 0, & x\leq 0.\end{cases} \tag{38}$$

The mean and variance of the exponential distribution are $1/\lambda$ and $1/\lambda^2$, respectively. Based on the assumption of the traffic demand, Eq. (34) can be obtained as follows:

$$E[\pi_p(b,D)]=-\frac{r+p}{\lambda}e^{-\lambda b}+\frac{r}{\lambda}-cb. \tag{39}$$

By using the definition of expectation and method of integral by parts, we can obtain the first term in Eq. (35) as follows:

$$\begin{aligned}E\left[(\pi_p(b,D))^2\right]&=e^{-\lambda b}\left(-\frac{2b}{\lambda}r^2-\frac{2}{\lambda^2}r^2+\frac{2}{\lambda^2}p^2+\frac{2pcb}{\lambda}+\frac{2rcb}{\lambda}\right)\\&-\frac{2rpb}{\lambda}e^{-2\lambda b}+\frac{2r^2}{\lambda^2}-\frac{2rcb}{\lambda}+c^2b^2.\end{aligned} \tag{40}$$

With the similar method, we can obtain the second term in Eq. (35) as follows:

$$\begin{aligned}\left(E[\pi_p(b,D)]\right)^2 &= e^{-\lambda b}\left(-\frac{2r^2-2rp}{\lambda^2}+\frac{2rcb+2pcb}{\lambda}\right)+\frac{r^2}{\lambda^2}\\&-\frac{r^2+2rp+p^2}{\lambda^2}e^{-2\lambda b}-\frac{2rcb}{\lambda}+c^2b^2.\end{aligned} \tag{41}$$

Substituting Eqs. (40) and (41) into Eq. (35), we can obtain that

$$\begin{aligned}Var[\pi_p(b,D)]&= e^{-2\lambda b}\left(-\frac{r^2+2rp+p^2}{\lambda^2}-\frac{2pcb}{\lambda}\right)\\&+e^{-\lambda b}\left(\frac{2rp}{\lambda^2}+\frac{2p^2}{\lambda^2}-\frac{2br^2}{\lambda}\right)+\frac{r^2}{\lambda^2}.\end{aligned} \tag{42}$$

The first order derivative of $Var[\pi_p(b,D)]$ of Eq. (42) with respect to b is given as follows:

$$\begin{aligned}\frac{\partial Var[\pi_p(b,D)]}{\partial b}&=2e^{-2\lambda b}\left(\frac{r^2+rp+p^2}{\lambda}+2rpb\right)\\&+2e^{-\lambda b}\left(br^2-\frac{r^2+rp+p^2}{\lambda}\right).\end{aligned} \tag{43}$$

Comparing with the results without the penalty cost presented in Section 4, where the variance function is an increasing function of the bandwidth capacity, there is no such relationship between $Var[\pi_p(b,D)]$ and b presented in Eq. (43) of this paper. The tradeoff between the mean function and the variance function becomes much more complicated when the penalty cost is introduced in this section than that presented in Section 4.

6 Numerical Results

In this section, based on the assumption of traffic load in a CN, we analyze the impacts of the relationship between unit revenue and unit cost, the loss rate constraint and the risk averseness on the system performance through numerical results, which are helpful for network traffic engineering design and planning.

We consider a fully distributed communication network, where the traffic demand offered to the whole CN forms a Poisson process with arrival rate $\lambda>0$, since in most of the system models in CNs, the arrival process of traffic demand can be assumed to form a Poisson process.

The interarrival times are exponentially distributed with rate λ. Let X be a random variable representing the time between successive demand arrivals in the Poisson process, then we have the probability distribution function $F_X(x)$ and the probability density function $f_X(x)$ of X presented in Eqs. (37) and (38). The mean and variance of the exponential distribution are $1/\lambda$ and $1/\lambda^2$, respectively.

Based on the assumption of the traffic demand, Eq. (2) can be obtained as follows:

$$E[\pi(b,D)] = -\frac{r}{\lambda}e^{-\lambda b} + \frac{r}{\lambda} - cb. \tag{44}$$

By using the definition of expectation and method of integral by parts, we can obtain the first term in Eq. (17) as follows:

$$E\left[(\pi(b,D))^2\right] = e^{-\lambda b}\left(-\frac{2b}{\lambda}r^2 - \frac{2}{\lambda^2}r^2 + \frac{2rcb}{\lambda}\right) + \frac{2r^2}{\lambda^2} - \frac{2rcb}{\lambda} + c^2b^2. \tag{45}$$

With the similar method, we can obtain the second term in Eq. (17) as follows:

$$\left(E[\pi(b,D)]\right)^2 = e^{-\lambda b}\left(-\frac{2r^2}{\lambda^2} + \frac{2rcb}{\lambda}\right) + \frac{r^2}{\lambda^2} - \frac{r^2}{\lambda^2}e^{-2\lambda b} - \frac{2rcb}{\lambda} + c^2b^2. \tag{46}$$

Substituting Eqs. (45) and (46) into Eq. (17), we can obtain that

$$Var[\pi(b,D)] = -\frac{r^2}{\lambda^2}e^{-2\lambda b} - \frac{2br^2}{\lambda}e^{-\lambda b} + \frac{r^2}{\lambda^2}. \tag{47}$$

In general, the maximal profit cannot always be obtained because of the loss rate constraint. In all numerical examples of this section, we let Δb denote the deviation of the bandwidth capacity from the optimal bandwidth without the loss rate constraint and increases from 0 to 10 by 1 each step.

In a practical CN, what a network manager concerns most is the profit, which is determined by the value of unit revenue r by serving the traffic demand and the unit cost c for unit bandwidth capacity allocated in the CN.

Without loss of generality, we assume that the value of unit revenue r by serving the traffic demand and the value of the unit cost c for unit bandwidth capacity allocated in the CN satisfy that $r = k \cdot c\,(k>0)$. In practical network management, if the value of k is set too large, the traffic demand will decrease, but if the value of k is set too small, the network manager side has no incentive to serve. Therefore, in Fig. 1 and Fig. 2, we show the impact of the relationship between r and c on the CN profit function, which gives insights for us to choose system parameters r and c in the following numerical analysis.

The horizontal axes of Figs. 1 and 2 correspond to the bandwidth deviation Δb from the optimal bandwidth b^*. The ordinate axes of Figs. 1 and 2 correspond to the percentage difference Π/Π^* from the benchmark Π^*, where Π^* is the optimal profit without the loss rate constraint given by Eq. (3), Π is the optimal profit with the loss rate constraint. Our numerical results include the optimal profit Π^* obtained without the loss rate constraint, which is one point with the ordinate axis value 100% and the horizontal axis value $\Delta b = 0$ of Figs. 1 and 2.

Figs. 1 and 2 show how change the percentage difference Π/Π^* with different values of $k = 2, 3, 5$ for the traffic rate $\lambda = 0.1$ and $\lambda = 0.3$, respectively.

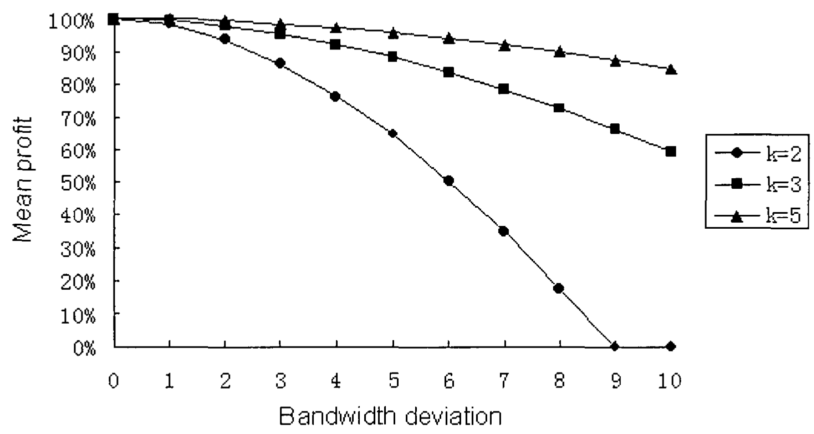

Figure 1: Impact of the value of k.

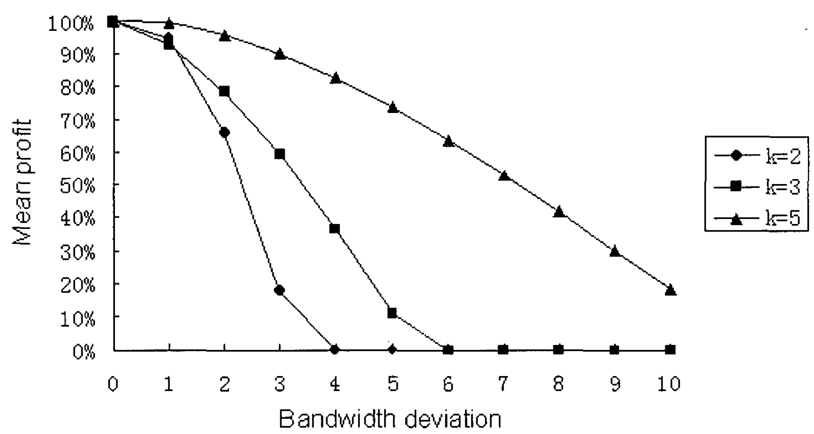

Figure 2: Impact of the value of k.

From the numerical results in Figs. 1 and 2, we can obtain that:

(1) The profit function in Fig. 1 decreases quickly when the value of k is smaller, i.e., $k=2$.

(2) The profit function in Fig. 2 decreases quickly when the value of k is smaller, i.e., $k=2,3$.

(3) Comparing the profit functions in both of Figs. 1 and 2, we note that the larger traffic rate λ is, the larger decrease of the profit function is.

In practical network, if the value of k has a great impact on system performance just as $k=2$ in Fig. 1 and $k=2,3$ in Fig. 2, then the system will not be stable. So, we should choose $k=5$ as the relationship between r and c to guarantee a stable CN for the numerical analysis.

According to the above analysis, we know that the system performance of CN is influenced by $k=5$ in nature. In all following numerical results, we choose system parameters $r=7.5$, $c=1.5$ by using the above analysis.

6.1 Impact of loss rate constraint on profit function

In this subsection, we give some numerical results to show the impact of loss rate constraint on the network profit function.

According to the engineering experience, we choose several different arrival rates to represent the different cases of traffic load in the CN as follows: $\lambda=0.01, 0.1, 0.5, 0.9$. Where $\lambda=0.01$ represents the case that the traffic load in the CN is low, $\lambda=0.1$ and $\lambda=0.5$ represent the cases that the traffic load in the CN is normal, and $\lambda=0.9$ represents the case that the traffic load in the CN is heavy.

The maximal profit cannot always be obtained because of the loss rate constraint. Let Δb denote the deviation of the bandwidth capacity from the optimal bandwidth without the loss rate constraint and increases from 0 to 10 by 1 each step with all other parameters unchanged.

The horizontal axis of Fig. 3 corresponds to the bandwidth deviation Δb from the optimal bandwidth b^*. The ordinate axis of Fig. 3 corresponds to the percentage difference Π/Π^* from the benchmark Π^*, where Π^* is the optimal profit without the loss rate constraint, Π is the optimal profit with the loss rate constraint.

Our numerical results include the optimal profit Π^* obtained without the loss rate constraint presented in [7], which is one point with the ordinate axis value 100% and the horizontal axis value $\Delta b=0$ of Fig. 3.

From the numerical results in Fig. 3, we can obtain that:

(1) In all curves, the optimal profit obtained with the loss rate constraint decreases as the deviation of bandwidth increases.

(2) The curve with a larger arrival rate has a quicker decreasing speed than the curve with a smaller arrival rate.

Compared with the results presented in [7] (without loss rate constraint), the numerical

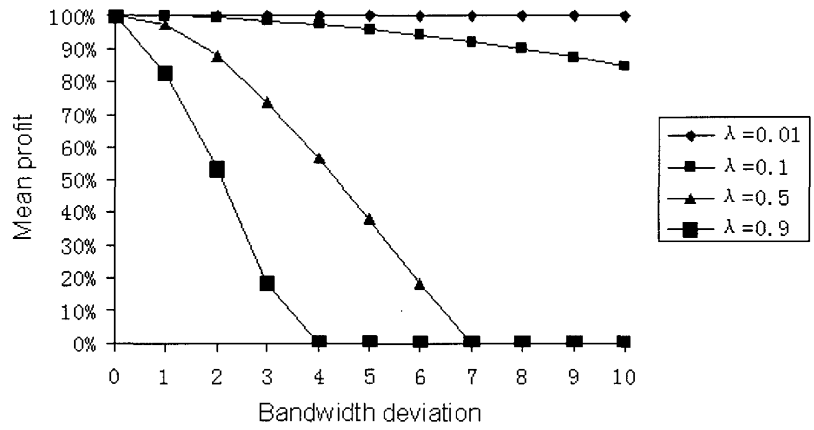

Figure 3: Impact of loss rate constraint on profit function.

results in our paper reveal a distinct impact of loss rate constraint on the network profit function.

6.2 Impact of risk averseness on bandwidth capacity

With the same system parameters for Fig. 3, we give some numerical results to show the impact of risk averseness on the network bandwidth capacity.

Let the risk averseness parameter α increase from 0.0 to 1.0 by 0.1 each step with all other parameters unchanged.

The horizontal axis of Fig. 4 corresponds to the risk parameter α taking values in [0.0, 1.0]. The ordinate axis of Fig. 4 corresponds to the percentage difference b/b^* from the benchmark b^*, where b^* is the optimal bandwidth obtained without the risk averseness, b is the optimal bandwidth obtained with the risk averseness.

Our numerical results include the optimal bandwidth b^* obtained without risk averseness, which is one point with the ordinate axis value 100% and the horizontal axis value $\alpha = 0.0$ of Fig. 4.

From the numerical results in Fig. 4, we can obtain that:

(1) In all curves, the optimal bandwidth obtained with the risk averseness decreases with the increase of risk averseness.
(2) The curve with a smaller arrival rate has a quicker decreasing speed than the curve with a larger arrival rate.
(3) All curves reveal a sharp decrease in [0.0, 0.1], and a tempered decrease in [0.1, 1.0].

Compared with the model without risk averseness, the numerical results in our paper reveal a distinct impact of risk averseness on the network bandwidth capacity.

6.3 Impact of risk averseness on objective function

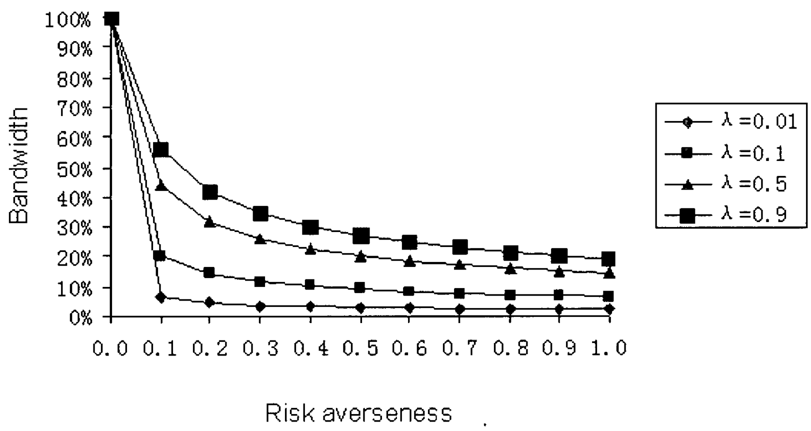

Figure 4: Impact of risk averseness on bandwidth capacity.

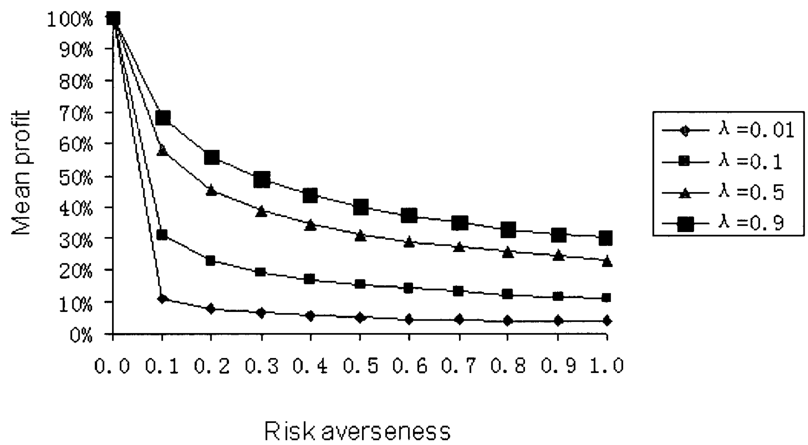

Figure 5: Impact of risk averseness on objective function.

With the same system parameters for Fig. 3, we give some numerical results to show the impact of risk averseness on the network objective function.

Let the risk averseness parameter α increase from 0.0 to 1.0 by 0.1 each step with all other parameters unchanged.

The horizontal axis of Fig. 5 corresponds to the risk parameter α taking values in [0.0, 1.0]. The ordinate axis of Fig. 5 corresponds to the percentage difference Φ/Φ^* from the benchmark Φ^*, where Φ^* is the optimal objective function obtained without the risk averseness, Φ is the optimal objective function obtained with the risk averseness.

Our numerical results include the optimal objective function Φ^* obtained without risk averseness, which is one point with the ordinate axis value 100% and the horizontal axis value $\alpha = 0.0$ of Fig. 5.

From the numerical results in Fig. 5, we can obtain that:

(1) In all curves, the optimal objective function obtained with the risk averseness decreases with the increase of risk averseness.
(2) The curve with a smaller arrival rate has a quicker decreasing speed than the curve with a larger arrival rate.
(3) All curves reveal a sharp decrease in [0.0, 0.2], and a tempered decrease in [0.2, 1.0].

Compared with the model without risk averseness, the numerical results in our paper reveal a distinct impact of risk averseness on the network objective function.

6.4 Impact of penalty cost on bandwidth capacity

With the same system parameters for Fig. 3, we give some numerical results to show the impact of penalty cost on the network bandwidth capacity.

Note that the optimal bandwidth capacity without penalty cost is $F^{-1}(\frac{r-c}{r})$. In this paper, the optimal bandwidth capacity with the penalty cost is given by Eq. (30).

We choose the unit revenue r as the benchmark of the penalty cost p. Let the percentage difference of the penalty cost increase from 0.0 to 1.0 by 0.1 each step with all other parameters unchanged.

The horizontal axis of Fig. 6 corresponds to the increase of the penalty cost p/r. The ordinate axis of Fig. 6 corresponds to the percentage difference of the optimal bandwidth capacity from the benchmark b^*, where b^* is the optimal bandwidth capacity without the penalty cost, and b is the optimal bandwidth capacity with the penalty cost.

Our numerical results include the optimal bandwidth capacity obtained without penalty, which is one point with the ordinate axis value 0% and the horizontal axis value $p/r = 0.0$ of Fig. 6.

From the numerical results shown in Fig. 6, we can conclude that:

(1) In all curves, the impact of the penalty cost on the bandwidth capacity increases as

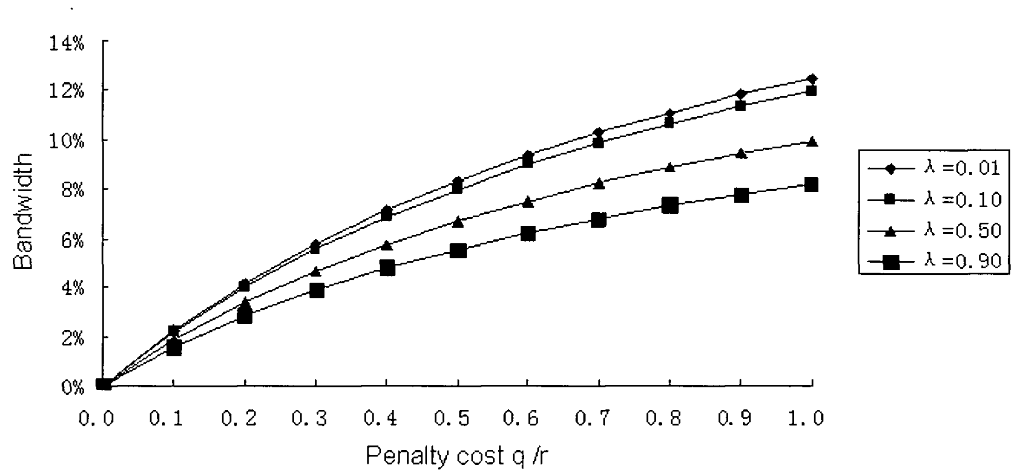

Figure 6: Impact of penalty cost on bandwidth capacity.

the penalty cost increases lineally.

(2) The curve with a smaller arrival rate has a quicker increasing speed than the curve with a larger arrival rate.

(3) With the same penalty cost, the heavier the traffic load in the CN is, the less the impact of the penalty cost on the bandwidth capacity will be.

Compared with the model without penalty cost, the numerical results in our paper reveal a distinct impact of penalty cost on the network bandwidth capacity. It implies that if we consider the penalty cost, the CN needs to be allocated more bandwidth capacity to guarantee the network performance.

6.5 Impact of penalty cost on profit function

With the same system parameters for Fig. 3, we give some numerical results to show the impact of penalty cost on the network profit function.

Note that the mean profit function $\Pi(b, D)$ without penalty cost is given as follows:

$$\Pi(b, D) = r\int_0^b xf(x)dx + rb\int_b^{+\infty} f(x)dx - cb.$$

In this paper, the the mean profit function with the penalty cost is given by Eq. (25).

We choose the unit revenue r as the benchmark of the penalty cost p. Let the percentage difference of the penalty cost increase from 0.0 to 1.0 by 0.1 each step with all other parameters unchanged.

The horizontal axes of Figs. 7 and 8 correspond to the increase of the penalty cost p/r. The ordinate axes of Figs. 7 and 8 correspond to the percentage difference of the optimal profit from the benchmark $\Pi_p^*(b, D)$, where $\Pi_p^*(b, D)$ is the optimal profit without the penalty cost, and $\Pi_p(b, D)$ is the optimal profit with the penalty cost.

Our numerical results include the optimal profit obtained without penalty, which are the points in the ordinate axes corresponding to $p/r = 0.0$ of Figs. 7 and 8.

From the numerical results shown in Figs. 7 and 8, we can conclude that:

(1) In all curves, the impact of the penalty cost on the mean profit function increases as the penalty cost increases lineally.

(2) The curve with a smaller arrival rate has a quicker decreasing speed than the curve with a larger arrival rate.

(3) With the same penalty cost, the heavier the traffic load in the CN is, the less the mean profit will be.

Compared with the model without penalty cost, the numerical results in our paper reveal a distinct impact of the penalty cost on the network profit function. Moreover, the numerical results with different arrival rates almost have the same increasing speed and impact on the profit function.

7 Conclusions

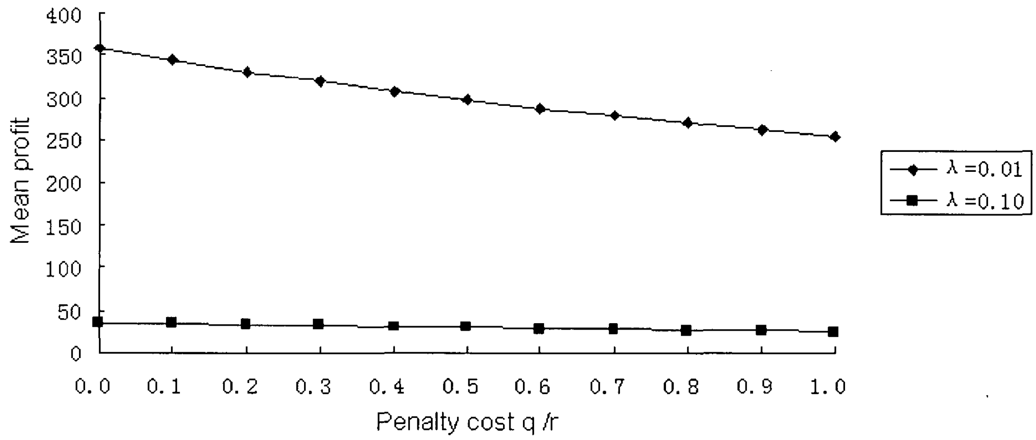

Figure 7: Impact of penalty cost on profit function.

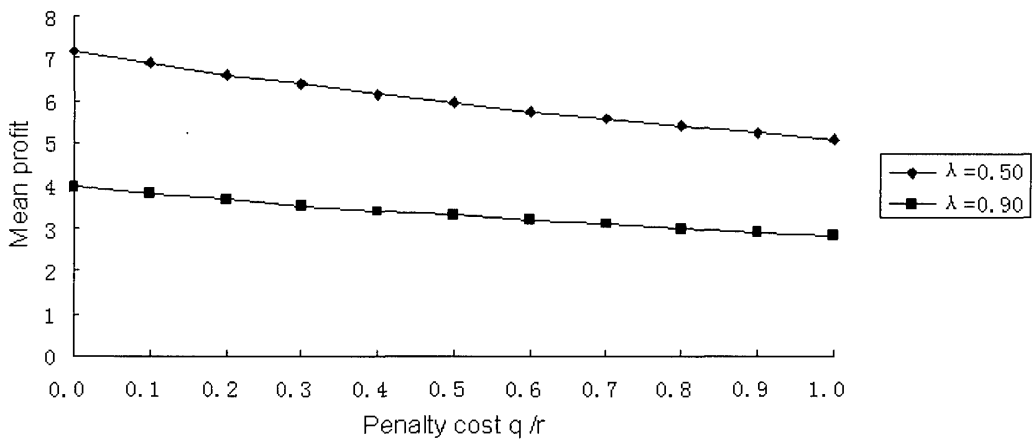

Figure 8: Impact of penalty cost on profit function.

In this paper, we presented a stochastic model for optimizing bandwidth allocation in Communication Networks (CNs) with a loss rate constraint, a profit shortfall risk and a linear penalty cost. The model was proposed for offline traffic engineering optimization, taking a centralized view of bandwidth allocation, performance control and risk of profit shortfall. We derived the optimal bandwidth allocation capacities with loss rate constraint and penalty cost, respectively. We used mean-variance approach to analyze the risk averseness in the CNs. We gave numerical results to compare our model with the model presented in [7] and revealed the impacts of loss rate constraint, risk averseness and penalty cost on the network performance.

We can conclude that the loss rate constraint, risk averseness and penalty cost have distinct impacts on the network performance. The implications presented in this paper reveal insights for traffic engineering design and planning in the CNs.

Acknowledgment

This work was supported in part by GRANT-IN-AID FOR SCIENTIFIC RESEARCH (No. 16560350) and by MEXT.ORC (2004-2008), Japan and in part by NSFC and MADIS, China.

References

[1] D. Awduche, A. Chiu, A. Elwalid, I. Widjaja and X. Xiao, "Overview and principles of internet traffic engineering," *RFC 3272, IETF*, 2002.

[2] X. Xiao, A. Hannan, B. Bailey and L. M. Ni, "Traffic engineering with MPLS in the Internet," *IEEE Network*, Vol. 14, pp. 28-33, 2000.

[3] P. Aukia et al., "RATES: a server for MPLS traffic engineering," *IEEE Network*, Vol. 14, pp. 34-41, 2000.

[4] A. Elwalid, C. Jin, S. Low, and I. Widjaja, "Mate: MPLS adaptive traffic engineering," *Proceedings of IEEE INFOCOM*, pp. 1300-1309, 2001.

[5] D. Mitra and K. G. Ramakrishnan, "A case study of multiservice multipriority traffic engineering design for data networks," *Proceedings of IEEE GLOBECOM*, pp. 1077-1083, 1999.

[6] S. Suri, M. Waldvogel, D. Bauer and P. R. Warkhede, "Profile-based routing and traffic engineering," *Computer Communications*, Vol. 26, pp. 351-365, 2003.

[7] D. Mitra and Q. Wang, "Stochastic traffic engineering, with applications to network revenue management," *Proceedings of IEEE INFOCOM*, pp. 1300-1309, 2001.

[8] D. Mitra and Q. Wang, "Risk-aware network profit management in a two-tier market," *Proceedings of 18th International Telegraffic Congress*, 2003.

[9] D. Mitra and Q. Wang, "Stochastic traffic engineering for demand uncertainty and risk-aware network revenue management," *Performance Evaluation Review*, Vol. 32, pp. 1-1, 2004.

[10] H. M. Markowitz, *Portfolio Selection, Efficient Diversification of Investments*, Yale University Press, New Haven, 1959.

[11] S. Wang and Y. Xia, *Portfolio Selection and Asset Pricing*, Springer-Verlag, Berlin, 2002.